ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/263260425

How to Pass a Turing Test

Article *in* Journal of Logic Language and Information · January 2000

DOI: 10.1007/978-94-010-0105-2_9

CITATIONS
17

READS
161

1 author:

William Rapaport
University at Buffalo, State University of New York
152 PUBLICATIONS **2,466** CITATIONS

SEE PROFILE

All content following this page was uploaded by William Rapaport on 29 March 2021.

The user has requested enhancement of the downloaded file.

How to Pass a Turing Test: Syntactic Semantics, Natural-Language Understanding, and First-Person Cognition
Author(s): William J. Rapaport
Source: *Journal of Logic, Language, and Information*, Vol. 9, No. 4, Special Issue on Alan Turing and Artificial Intelligence (Oct., 2000), pp. 467-490
Published by: Springer
Stable URL: http://www.jstor.org/stable/40180238
Accessed: 10/02/2010 15:23

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at http://www.jstor.org/page/info/about/policies/terms.jsp. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at http://www.jstor.org/action/showPublisher?publisherCode=springer.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Springer is collaborating with JSTOR to digitize, preserve and extend access to *Journal of Logic, Language, and Information.*

http://www.jstor.org

Journal of Logic, Language, and Information **9**: 467–490, 2000.
© 2000 *Kluwer Academic Publishers. Printed in the Netherlands.*

How to Pass a Turing Test

Syntactic Semantics, Natural-Language Understanding, and First-Person Cognition

WILLIAM J. RAPAPORT
Department of Computer Science and Engineering, Department of Philosophy, and Center for Cognitive Science, State University of New York at Buffalo, Buffalo, NY 14260-2000, U.S.A.
E-mail: rapaport@cse.buffalo.edu; http://www.cse.buffalo.edu/~rapaport

(Received 1 June 1999; in final form 15 April 2000)

Abstract. I advocate a theory of "syntactic semantics" as a way of understanding how computers can think (and how the Chinese-Room-Argument objection to the Turing Test can be overcome): (1) Semantics, considered as the study of relations *between* symbols and meanings, can be turned into syntax – a study of relations *among* symbols (including meanings) – and hence syntax (i.e., symbol manipulation) can suffice for the semantical enterprise (contra Searle). (2) Semantics, considered as the process of understanding one domain (by modeling it) in terms of another, can be viewed recursively: The base case of semantic understanding – understanding a domain in terms of itself – is "syntactic understanding." (3) An internal (or "narrow"), first-person point of view makes an external (or "wide"), third-person point of view otiose for purposes of understanding cognition.

Key words: Chinese-Room Argument, first-person point of view, internalism, methodological solipsism, problem of other minds, representative realism, rules and representations, semantic network, semantics, SNePS, syntax, Turing Test

> We now and then take pen in hand
> And make some marks on empty paper.
> Just what they say, all understand.
> It is a game with rules that matter.
>
> Hermann Hesse, "Alphabet,"
> trans. R.S. Ellis (Manin, 1977: 3)

1. The Turing Test

Turing opened his essay "Computing Machinery and Intelligence" by saying that he would "consider the question, 'Can machines think?' " (Turing, 1950: 433). Rather than answer this provocative question directly, he proposed his now-famous experiment, whose outcome would provide guidance on how to answer it. He described the experiment by analogy with a parlor game that he called "the 'imitation game' " (Turing, 1950: 433), in which an interrogator must decide which of two people of unknown gender is male (A) and which is female (B). He then asked,

> ... 'What will happen when a machine [specifically, a digital computer; p. 436] takes the part of A [the man] in this game?' Will the interrogator

> decide wrongly as often when the game is played like this as he does when the game is played between a man and a woman? These questions replace our original, 'Can machines think?' (Turing, 1950: 433–434).

Turing says nothing about what the suitably-programmed computer is supposed to do. Clearly, it is supposed to play the role of the man, but the man's task in the original imitation game was to fool the interrogator into thinking that he or she is conversing with the woman. Traditionally, this has been taken to mean that the computer is supposed to fool the interrogator into thinking that it is human *simpliciter*. However, read literally and conservatively, if the computer is supposed to do this by playing the role of the man, then it appears that the computer has a more complex task, namely, to behave like a man who is trying to convince the interrogator that he is a woman! (Colby et al., 1972: 202 make a similar observation.) Of course, were the computer to be successful in this very much harder task, it would also, *ipso facto*, be successful in convincing the interrogator that it was human *simpliciter*.

Later (p. 442), Turing considers "one particular digital computer *C*," and asks whether "*C* can be made to play satisfactorily the part of A [i.e., the man] in the imitation game, the part of B [i.e., the woman] being taken by a man?" If the part of B is taken by a man, then it follows, from the earlier description that the interrogator's task is to determine which of X and Y is A and B, that B is simply supposed to convince the interrogator that he is the man (or the human) and that the computer's task is to convince the interrogator that *it* is the man (or the human). So it appears that Turing was not overly concerned with the complication discussed in the previous paragraph (although he apparently thought it important that the human in this human-computer contest be represented by a man, not a woman). In any case, Turing answered this new question as follows:

> I believe that in about fifty years' time [i.e., by about 2000] it will be possible to programme computers ... to make them play the imitation game so well that an average interrogator will not have more than 70 per cent. chance of making the right identification after five minutes of questioning. The original question, 'Can machines think?' I believe to be too meaningless to deserve discussion. Nevertheless I believe that at the end of the century *the use of words* and *general educated opinion* will have altered so much that one will be able to speak of machines thinking without expecting to be contradicted (Turing, 1950: 442; my italics).

2. The Use of Words vs. General Educated Opinion

2.1. "THINKING" VS. THINKING

The Turing Test, as the computer version of the imitation game has come to be called, is now generally simplified even further to a 2-player game: Can a human conversing with an unknown interlocutor through a computer "chat" interface determine whether the interlocutor is a human or a suitably programmed computer, or – more simply – can a computer convince an interrogator (who is unaware of who or what he or she is conversing with) that its ability to think, as demonstrated by its ability to converse in natural language, is equivalent to that of a human (modulo the – quite low – 70%/5-minute threshold)?

There is an echo of this in Steiner's famous *New Yorker* cartoon (5 July 1993: 61) in which a dog, sitting in front of a computer, observes that "On the Internet, nobody knows you're a dog." The success of this cartoon depends on our realization that, in fact – just like the interrogator in a 2-player Turing test – one does *not* know with whom one is communicating over the Internet. This ignorance on our part can have serious real-life implications concerning, e.g., computer security (if I enter my credit-card number on a Web site, have I really bought a book, or have I given my number to a con artist?) and matters of social welfare or personal safety – even life and death (is my daughter chatting with a member of the opposite sex who is about her age, or with a potential sex offender?). But note also that, even though many of us are aware of these possibilities, we normally assume that we are *not* talking to a con artist, a sex offender, or even a dog. Or – for that matter – a computer.* (My mother did not recognize (or expect) the possibility that she was not talking to a human on the phone, and thus regularly tried to converse with pre-recorded phone messages.) We normally are, in fact, fully prepared to accept our invisible interlocutor as a (normal, ordinary) human with human thinking capacities.

And this, I suggest, was Turing's point.** It is, nearly enough, the point of the argument from analogy as a solution to the problem of other minds: I know (or assume) that *I* have a mind and can think, but, when I converse with you face to face, how do I know whether (or can I assume that) *you* have a mind and can think? The argument from analogy answers as follows: you are sufficiently like me in all other visible respects, so I can justifiably infer (or assume) that you are like me in this invisible one. Of course, I could be wrong; such is the nature of inductive inference: You could be a well-designed android whose natural-language-processing

* As I write this, researchers are beginning to investigate just such assumptions; see Berman and Bruckman (1999) and Hafner (1999).

** Although, as my colleague Stuart C. Shapiro pointed out to me, the interrogator, on one reading, knows that he or she is participating in a test. However, another Turing test does not require such knowledge on the interrogator's part: let the interrogator (unknowingly, of course) begin the conversation with the human; then, at some point, let the computer change places with the human. Can the interrogator tell at what point in the conversation the switch took place? (This is suggested in a passage in Lassègue (1996: §3.2.2). A similar suggestion ("the Extended Turing Test") was made in Abelson (1968: 317–320) and is discussed in Colby et al. (1972: 203–204).)

component is just an elaboration of Weizenbaum's (1966) Eliza program (cf. Section 8, below).* But we make this inference-to-mindedness – if only unconsciously – on a daily basis, in our everyday interactions. Now, in the case of a Turing test, I (as interrogator) have considerably less analogical information about you; I only have our conversations to go by. But, even in a much weaker case such as this, we do ordinarily infer or assume (and justifiably so) that our interlocutor is human, with human cognitive capabilities.

Is there anything wrong with this? Well, if my interlocutor is not who (or what) I think he (or she, or it) is, then I was wrong in my inference or assumption. And if my interlocutor was really a suitably programmed computer, then I was certainly wrong about my interlocutor's *biological* humanity. But was I wrong about my interlocutor's (human) cognitive capabilities (independently of the interlocutor's implementation)? That is the question. Turing's answer is: No. Perhaps more cautiously, the lesson of Turing's test is that the answer depends on how you define "(human) cognitive capabilities." One way to define them is in terms of "passing" a Turing test; in that case, of course, any Turing-test-passing interlocutor does think (this is essentially Turing's strategy). Another way is to come up with an antecedently acceptable definition, and ask whether our Turing-test-passing interlocutor's behavior satisfies it. If it does, we have several choices: (1) we could say that, therefore, the interlocutor does think, whether or not it is biologically human (this is, roughly, Turing's strategy, where the antecedently-given definition is something like this: convincing the interrogator of your cognitive capacities with the same degree of accuracy as, in the original game, the man (A) convinces the interrogator that he is the woman (B)); or (2) we could say that there must have been something wrong with our definition if the interlocutor is not biologically human; or (3) we could say that, while the interlocutor is doing something that superficially satisfies the definition, it is not "really" thinking. In case (3), we could go on to say (4) that that is the end of the matter (this is essentially Searle's move in the Chinese-Room Argument) or (5) that the interlocutor is merely "thinking" in some metaphorical or extended sense of that term. Comparison with two other terms will prove enlightening.

2.2. "FLYING" VS. FLYING

Do birds fly? Of course. Do people fly? Of course not, at least not in the same sense. When I say that I flew to New York City, I don't really mean that I flew like a bird. ("Didn't your arms get tired?," joke my literalistic friends.) I mean that I was a passenger on an airplane that flew to New York City. Oh? *Do* airplanes fly? Well, of course; don't they? Isn't that what the history of heavier-than-air flight was all about? Ah, but planes don't fly the way birds do: They don't flap their wings, and they are powered by fossil fuel. So have we, after all, failed in our centuries-old attempt to fly like the birds? No. But how can this be?

* I am indebted to an anonymous reviewer for a suggestion along these lines.

There are two ways in which it makes perfectly good sense to say that planes fly. One way is to say that 'fly' is used metaphorically with respect to planes – birds fly; planes only "fly" – but this is one of those metaphors that have become so ingrained in our everyday language that we no longer recognize them as such. Turing may have had this in mind when he spoke – in the italicized passage quoted above – about "the use of words" changing. Thus, we can likewise extend 'flying' to cover hot-air balloons (which do not have wings at all), spaceships (which do not travel in air), arrows and missiles (some of which, perhaps more accurately, merely "fall with style," as the film *Toy Story* puts it), and even the movement of penguins under water (more usually called 'swimming:' "But penguins do indeed fly – they fly in water. Using their wings, which are flat and tapered and have a rounded leading edge, and flapping like any swift or lark, penguins fly through the water to feed and to escape predators" (Ackerman, 1989: 45)).

The other way in which it makes perfectly good sense to say that planes fly is to note that, in fact, the physics of flight is the same for both birds and planes (e.g., shape of wing, dynamics of airflow, etc.). What we may have once thought was essential to flying – flapping of wings – turns out to be accidental. Our understanding of what flying really is has changed (has become more general, or more abstract), so that more phenomena come under the rubric of 'flying.' Turing may have had this option in mind in his remark about "general educated opinion" changing.

The same two options apply to 'thinking:' we could say that, insofar as suitably programmed computers pass a Turing test, they do think – extending "think" metaphorically, but legitimately, just as we have extended 'fly' (which we have always done, even at the very beginnings, centuries ago, of research into human flight). Or we could say that being human is inessential for thinking, the general nature of thinking being the same for both humans and suitably programmed computers (as well as animals).

In fact, *both* the use of the word 'fly' *and* general educated opinion have changed. Thus, some things (spaceships, missiles) arguably only "fly," while others (planes) definitely fly like birds fly. But one can in fact speak of all those things flying "without expecting to be contradicted." Moreover, these two ways need not be exclusive; the common physical or psychological underpinnings of flight or thought might be precisely what allow for the seamless metaphorical extension.

2.3. "COMPUTER" VS. COMPUTER

Another term that has undergone a change of meaning is also instructive and perhaps more to the point: 'computer.'* At the time of Turing's 1936 paper on what is now called the Turing machine, a "computer" was primarily a *human* who

* I am grateful to Alistair E. Campbell for this example.

computed.* Turing distinguished between a computing *machine* and a (human) *computer*:

> The behaviour of the *computer* at any moment is determined by the symbols which *he* is observing, and *his* 'state of mind' at that moment. ... We may now construct a *machine* to do the work of this computer. To each state of mind of the *computer* corresponds an '*m*-configuration' of the *machine* (Turing, 1936 [1965: 136–137]; my italics).

By the time of his 1950 paper, he posed the question "Can *machines* think?" and spoke of "*digital* computers," "*electronic* computers," and "*human* computers," only rarely using 'computer' unmodified to mean a computing *machine*, as if the modifier 'digital' or 'electronic' still served to warn some readers that *human* computers were not the topic of discussion. Today, "computer" almost never refers to a human.

What happened here? Perhaps first by analogy or metaphorical extension, 'computer' came to be applied to machines. And then, over the years, it has been applied to a large variety of machines: vacuum-tube computers, transistor-based computers, VLSI computers, mainframes, workstations, laptops, "Wintel" machines, Macs, even special-purpose microprocessors embedded in our cars, etc. What do all these (as well as humans) have in common? – the ability to compute (in, say, the Turing-machine sense). Thus, "general educated opinion" has changed to view 'computer,' not so much in terms of an implementing device, but more in terms of functionality – input-output behavior, perhaps together with general algorithmic structure. This change in 'computer' to focus on computational essentials parallels the change in 'fly' to focus on aerodynamic essentials. And it parallels a change in 'think' (and its cognates) to focus on the computational/cognitive essentials. So it is quite possible that Turing was suggesting that the use of 'think' (and its cognates) will undergo a similar conversion from applying only to humans to applying also (albeit not primarily) to machines.

"But," the critic objects, "it isn't *really* thinking; there's more to thinking than passing a Turing test." This is the gut feeling at the heart of Searle's (1980) Chinese-Room Argument, to which we now turn.

3. The Chinese-Room Argument

The Chinese-Room Argument sets up a situation in which an entity passes a Turing test but, by hypothesis, cannot "think" – more specifically, cannot understand language. In this section, I present the argument and two objections.

---

* In the *OED* (Simpson and Weiner, 1989: Vol. III, 640–641), the earliest cited occurrence of 'computer' (1646) refers to humans. The earliest citation for 'computer' referring to machines is 1897, and the next is 1915, both long before the development of modern computers; the bulk of the citations are from 1941ff.

3.1. The Argument

The situation is this: Searle, who by hypothesis cannot understand written or spoken Chinese, is sealed in a room supplied with paper, pencils, and an instruction book written in English (which he does understand). (1) Through an input slot come pieces of paper with various marks ("squiggles") on them. (2) Searle-in-the-room manipulates the squiggles according to the instructions in the book, and outputs other pieces of paper with squiggles on them that he wrote following the instructions. Steps (1) and (2) are repeated until the experiment stops. From Searle-in-the-room's point of view, that is all he is doing. Unknown to him, however, outside the room (playing the role of interrogator in a Turing test) is a native speaker of Chinese. This native speaker has been inputting to the room pieces of paper with a story (written in Chinese), sufficient background information (written in Chinese) for whoever (or whatever) is in the room to understand the story, and questions (written in Chinese) about the story. And the native speaker has been receiving, as output from the room (or from whoever or whatever is in it), pieces of paper with excellent answers to the questions, written in fluent Chinese. From the native speaker's point of view, whoever or whatever is in the room understands Chinese and thus has passed this Turing test (but see Section 6.1, below, on the accuracy of this description). But the native speaker's and Searle-in-the-room's points of view are inconsistent; moreover, Searle-in-the-room's point of view is, by hypothesis, the correct one. Therefore, it is possible for an entity to pass a Turing test without being able to think. More precisely, it is possible to pass a Turing test for understanding natural language without being able to understand natural language. (I return to the differences in point of view in Section 6.1, below.)

3.2. Two Objections

There have been numerous objections to the Chinese-Room Argument right from the beginning (cf. Searle, 1980), but this is not the place to survey them all I will focus on only two of them. At its core, there are two components to "the" Chinese-Room Argument: an argument from biology and an argument from semantics.

3.2.1. *The Argument from Biology*

The argument from biology is this:

(B1) Computer programs are non-biological.
(B2) Cognition is biological.
(B3) So, no non-biological computer program can exhibit cognition.

I claim that (B2) is wrong: it assumes that cognition (in particular, understanding natural language) is not something that can be characterized abstractly and implemented in different (including non-biological) media (cf. Rapaport, 1985, 1986a,

1988a, 1996: Ch. 7, 1999). But if – and I readily admit that this is a big "if" – computational cognitive science succeeds in its goal of developing an algorithmic theory of cognition, then those algorithms will be implementable in a variety of media, including non-biological ones.* And any medium that implements those algorithms will exhibit cognition (just as airplanes, as well as birds, do fly). (For a defense of this against two recent objections, see Rapaport (1998).)

3.2.2. *The Argument from Semantics*

The present essay is concerned with the argument from semantics:

(S1) Computer programs are purely syntactic.
(S2) Cognition is semantic.
(S3) Syntax alone is not sufficient for semantics.
(S4) So, no purely syntactic computer program can exhibit semantic cognition.

I claim that premise (S3) is wrong: Syntax *is* sufficient for semantics. Now, anyone who knows what "syntax" and "semantics" are knows that they are not the same thing – indeed, I spend hours each semester trying to drive home to my students what the differences are. So how can I turn around and say that one suffices for the other?

To begin to see how, consider that what Searle alleges is missing from the Chinese Room is semantic links to the external world, links of the form that such-and-such a squiggle refers to, say, hamburgers: "... I still don't understand a word of Chinese and neither does any other digital computer because all the computer has is what I have: a formal program that attaches no meaning, interpretation, or content to any of the symbols" (Searle, 1982: 5). Note that Searle makes two assumptions: that external links are needed for the program to "attach" meaning to its symbols, *and* a solipsistic assumption that the computer has no links to the external world – that all is internal to it. Now, first, *if* external links *are* needed, then surely a computer could have them as well as – and presumably in the same way that – humans have them (this, I take it, is the thrust of the "robot" reply to the Chinese-Room Argument; Searle, 1980: 420). But *are* external links needed? How might we provide Searle-in-the-room with such links? One way would be to give him, say, a hamburger (i.e., to import it from the external world) clearly labeled with the appropriate squiggle. But now the hamburger is *in* the room; it is no longer part of the external world. Sure – it *came* from the external world, but so did the squiggles. Searle-in-the-room could just as well have been antecedently supplied with a stock of sample objects (and much else besides, for word-*object* links will not suffice;

* Conceivably, some of the algorithms might be implementation-dependent in some way; see, e.g., Thagard (1986); cf. Maloney (1987). But at most this might constrain the nature of the feasible implementing media. It would not necessarily rule out non-biological ones. In any case, the view that an algorithm might be implementation-dependent would seem to go against the grain of the generally accepted view of algorithms as being implementation-*in*dependent.

abstract concepts such as *love*, *number*, etc., will require word-*concept* links).* In either case (an imported hamburger delivered from outside or a previously-supplied one stored in the refrigerator at home), the word-meaning links would be internal to the room. As I will argue below, this makes them part of a (larger) syntactic system, and so syntax will *have* to suffice for semantics. To see how, it will help if we review the classical theory of syntax and semantics.

4. Syntax and Semantics: Games with Rules

Consider some symbol system, i.e., some set of symbols that may or may not be "meaningful." Now, I am stepping on some semiotic toes here when I talk like this, for, in the vocabulary of many (if not most) writers on the subject, symbols are, by definition, meaningful. So, instead, consider a set of "markers" (let us call them) that do not wear any meaning on their sleeves (cf. Fetzer, 1994: 14; Rapaport, 1998). Think of marks or patterns on paper (or some other medium) that are easily re-identifiable, distinguishable one from another, relatively unchanging, and do not (necessarily) come already equipped with a semantic interpretation.

According to Morris's classic presentation of semiotics (1938: 6–7), *syntax* is the study of relations among these markers. Some, for instance, are proper parts of others; certain combinations of them are "legal" (or "grammatical"), others not; and whenever some are in proximity to each other, certain others can be constructed or "derived" from them; etc. (This characterization is intended to cover both the well-formedness rules of complex markers as well as proof-theoretical rules of inference.) Crucially, syntax does not comprise any relations of the markers to any non-markers.

Semantics, according to Morris, is precisely what syntax is not: the study of relations *between* the system of markers and other things. What other things? Traditionally, their "meanings:" Traditionally, semantics is the study of the relation of symbols to the things (in the world) that the symbols mean.**

What is not usually noticed in these definitions is this: if the set of markers is unioned with the set of meanings,‡ and the resulting set considered as a set of (new)

* Cf. Swift (1726: Pt. III, Ch. 5) [1967: 230f]. Moreover, as Kearns (1997) has argued, it is *speech acts*, not expressions, that are the bearers of meaning.

** *Pragmatics* will be of less concern to us, but, for the sake of completeness, let me mention that pragmatics is, according to Morris, the study of the relations between markers and their interpreters. Note that this tripartite analysis of semiotics omits a study of the relations between interpreters and symbol-meanings, as well as studies of the relations among symbol-meanings (or is that all of science and perhaps some of psychology?) and of the relations among interpreters (or is that part of sociology?). Perhaps as a consequence, pragmatics is often described as the study of the relations among markers, their meanings, *and* users of the markers. This somewhat more vague study has variously been taken to include the study of indexicals (symbols whose meaning depends on speaker and context), speech acts, discourse phenomena, etc.; it is often characterized as a grab bag of everything not covered by syntax and semantics as above defined.

‡ Taking care in the case of markers that refer to other markers as their meanings, an important special case that I want to ignore for now.

markers (i.e., if the "meanings" are made internal to the symbol system), then *what was once semantics* – viz., relations *between* old markers and their meanings – *is now syntax* – viz., relations *among* old and new markers (see Section 5; these new relations are in *addition* to the old ones that classify markers and provide well-formedness rules). Furthermore, it is left open how the symbol-user understands the symbol-meanings (see note about pragmatics above). I shall argue that this must be done syntactically (Section 7). It is in these ways that syntax can suffice for semantics.

But a lot more needs to be said.

5. Syntactic Semantics: I – Turning Semantics into Syntax

One thing that is needed is an argument that the set of (old) markers *can* be unioned with the set of meanings. Insofar as the markers are internal to a mind, we need an argument that the semantic domain can be internalized, so to speak. This *can* happen under certain conditions. In particular, it happens under the conditions obtaining for human language understanding. For how *do* I learn the meaning of a word? Let us, for now, consider only the very simplest case of a word that clearly refers.* How do I learn that 'tree' refers to that large brown-and-green thing I see before me? Someone points to it in my presence and says something like "This is called a 'tree'." Perhaps numerous repetitions of this, with different trees, are needed. I begin to associate** two things, but what two things? A tree and the word 'tree'? No; to paraphrase Percy (1975: 43), the tree is not the tree out there, and the word 'tree' is not the sound in the air.‡ Rather, *my internal representation of* the word becomes associated ("linked," or "bound") with *my internal representation of* the tree.‡‡ Light waves reflected from the tree in the external world enter my eyes, are focused on my retina, and are transduced into electrochemical signals that travel along my optic nerve to my visual cortex. No one knows *exactly* what goes on in visual cortex (or elsewhere) at that point. But surely some nerves are activated that are my internal representation (perhaps permanent, perhaps fleeting) of that

* The case of other terms is even more likely to be internal; this is best explored in the context of conceptual-role semantics; cf. Rapaport (1996: Ch. 4).

** What constitutes "association"? In this case, simply co-occurrence: when I hear 'tree,' I think of trees. Later, it will mean that some kind of "internal" link is forged between the associated things: in the case of Cassie, a computational cognitive agent (introduced later in this section), it will be a semantic-network path; in the case of a human, it might be some neural sort of "binding" (see Damasio, 1989).

‡ Although apt, this is a slightly misleading paraphrase, since Percy's point is that, in understanding that 'tree' means tree, 'tree' and tree are types, not tokens.

‡‡ In the case of Cassie, an "internal representation" of a word or object would be a semantic-network node. In the case of a human, it might be a pattern of neuron firings. 'Representation' may not be the happiest term: if there is an external object, then the internal correlate "represents" it in the sense that the internal entity is a proxy for the external one. But if there is no external entity (as in the case of 'unicorn'), then it is perhaps inappropriate to speak of 'representation.' See Rapaport (1978, 1981) and Shapiro and Rapaport (1987, 1991) for more on the nature of these Meinongian objects.

tree. Likewise, sound waves emanating from the 'tree'-speaker's vocal tract reach my ears and, via my auditory nerves, ultimately reach my auditory cortical areas, where surely the story is the same: some nerves are activated that are my internal representation (for the nonce, if not forever) of the word 'tree.' *And these two sets of activated nerves are, somehow, associated, or "bound."* (For some discussion of this, see Damasio (1989) and Rapaport (1996: esp. Ch. 3).) *That* is the semantic relation, but – taking the activated nerves as the markers (as well as the meanings) – it is a *syntactic* relation. (Here, 'syntax,' qua "symbol manipulation" (or "marker manipulation"), is to be taken broadly. For discussion, see Bunn (forthcoming) and Jackendoff (forthcoming).) Thus, it is precisely this coordination of multiple modalities that allows syntax to give rise to semantics.

The same holds – or could hold – for a suitably programmed computer. When I converse in English with "Cassie" – a computational cognitive agent implemented in the SNePS knowledge-representation, reasoning, and acting system – she builds internal representations (nodes of a semantic network) of my sentences (Shapiro, 1979; Shapiro and Rapaport, 1987, 1991, 1992, 1995). If I show her pictures, she builds similar internal representations (more nodes of the same semantic network), and she can associate the nodes from the "linguistic part" of her network with the nodes from the "visual part" (Srihari and Rapaport, 1989; Srihari, 1991). (The inverse task, of finding – or pointing to – some object in the external world, supplements the nodes with other symbols, as described in detail in Shapiro (1998); roughly, Cassie's internal representation of the object is "aligned" with, again roughly, her internal visual representation of it, and that latter symbol is used to direct her to the corresponding external entity, but in no case does she have direct access to the external entity.)*

6. Points of View

6.1. Whose Point of View Is "Correct"?

The internal-picture sketched above is from the first-person point of view. In studying how a human mind understands language (or, more generally, thinks and cognizes), and in constructing computational models of this (or, more strongly, constructing computational cognitive agents), we must consider, primarily, what

* A related argument for an apparently similar conclusion, based on Chomsky's "minimalist" program, has been offered in McGilvray (1998): "one should look ... to expressions inside the head for meanings .... [M]eanings are contents intrinsic to expressions ... and ... they are defined and individuated by syntax, broadly conceived. ... [T]hese concepts are individuated by internally and innately specified features, not by their relationships to the world, if any" (pp. 225, 228).

My merger of syntax and semantics into a new syntactic domain whose relation between old (syntactic) markers and new (semantic) markers seems to be echoed by Chomsky's " 'Relation *R*' ('for which read reference,' but without the idea that reference relates an LF [logical form] to something 'out there') that stands between elements of an LF and these stipulated semantic values that serve to 'interpret' it. This relation places *both* terms of Relation *R*, LFs and their semantic values, entirely within the domain of syntax, broadly conceived; .... They are in the head" (p. 268).

is going on inside the agent's head, from the agent's point of view. (In Chomsky's terms, we must study an "I-language"; cf. McGilvray (1998: 240–241).) Internally, there are markers that represent or correspond to linguistic entities (words, sentences), markers that represent or correspond to conceptual entities (e.g., propositions and their components), and (perhaps) markers that represent or correspond to entities in the external world.* But all of these internal markers are only related to each other. More precisely, the cognitive agent only needs to deal with (i.e., to manipulate) these internal markers; the agent does not need to be concerned with the causal-historical origins of the markers, nor do we need to be concerned with these origins insofar as we are trying to understand how the agent thinks by means of these markers. We need only study the internal relations among them. We do not (at this stage) need to study any external relations between markers and external entities.

The notion of "point of view" is central to the Turing-Test-vs.-Chinese-Room debate, too. As we saw in Section 3.1, the point of view of the native Chinese speaker differs from the point of view of Searle-in-the-room. Which point of view should dominate? The Turing Test only talks about the point of view of the interrogator; so – contrary to Searle – what might "really" be going on in the external world (i.e., the point of view of Searle-in-the-room) is irrelevant to the Turing Test.

To get a feeling for why this is, consider the following conversation between Dorothy and Boq (a Munchkin) from *The Wizard of Oz*:

> When Boq saw her silver shoes** he said,
>
> "You must be a great sorceress."
>
> "Why?" asked the girl.
>
> "Because you wear silver shoes and have killed the wicked witch. Besides, you have white in your frock, and only witches and sorceresses wear white."
>
> "My dress is blue and white checked," said Dorothy, smoothing out the wrinkles in it.
>
> "It is kind of you to wear that," said Boq. "Blue is the color of the Munchkins, and white is the witch color; so we know you are a friendly witch."
>
> Dorothy did not know what to say to this, *for all the people seemed to think her a witch, and she knew very well she was only an ordinary little girl* who had come by the chance of a cyclone into a strange land (Baum, 1900: 34–35; my italics).

Is Dorothy a witch or not? From her point of view, she is not; but, from the point of view of Boq and the other Munchkins, she is. Dorothy knows herself not to be a witch, no? At least, she *believes* that she is not a witch, as *she* understands that

* But see Maida and Shapiro (1982), Shapiro and Rapaport (1991) and Section 6.2, below, for an argument *against* representing external, or "extensional," entities.

** A note for those only familiar with the 1939 movie version: The novel has silver shoes, not ruby slippers. And, to those only familiar with the 1939 movie version, shame on you! Baum's *Oz* books are full of wonderful philosophical observations.

term. But it is certainly possible for her to *believe* that she is not a witch, yet for her to *really be* a witch (in either her terms or the terms of the Munchkins). So, what counts as really being a witch? We must answer this from the point of view of what *Munchkins* take witches to be, for there are many theories of witchcraft, but only the Munchkin theory counts in the present context. The dispute is not about whether Dorothy is "really" a witch in some context-independent sense (from Dorothy's, or the reader's, point of view), but whether she is a witch in the Munchkin sense (from the Munchkin point of view). Boq cites her clothing and actions, which Dorothy admits to. In *Oz*, witches also perform magic, which Dorothy denies having done. But what counts as magic (again from the Munchkin point of view)? Standard magical things like disappearing and transforming one object into another, to be sure, but who is Dorothy (or me, for that matter) to say that, from the Munchkin point of view, her behavior and actions (such as suddenly dropping from the sky) are not included under what they consider to be "magical"? The Munchkin point of view trumps Dorothy's point of view with respect to what it means to be a witch in Munchkinland – they, not Dorothy, are the experts on criteria of their notion of witchcraft.*

The Chinese-Room situation is analogous. Does *Searle-in-the-room* understand Chinese or not? (Note that this is the question that Searle (1980) himself poses; more on this below.) From his point of view, he does not; but from the point of view of the native Chinese speaker, he does. Searle-in-the-room knows himself not to understand Chinese, no? (Certainly, that is what Searle (1980) claims.) At least, he *believes* that he does not understand Chinese, as *he* understands that term. But it is certainly possible for him to *believe* that he does not understand Chinese, yet for him to *really understand* Chinese (see the next paragraph). So, what counts as really understanding Chinese? We must answer this from the point of view of what native Chinese speakers take understanding Chinese to be. For a person might believe that he or she *does* understand Chinese, yet be mistaken; only the native Chinese speaker can ask appropriate questions to determine whether that person really does understand. The native Chinese speaker's point of view trumps Searle-in-the-room's point of view with respect to what it means to understand Chinese – the native Chinese speaker, not Searle-in-the-room, is the expert on criteria of understanding Chinese.

The Chinese-Room case may need a bit more explication, for Searle-in-the-room could legitimately reply to the native Chinese speaker that he, Searle-in-the-room, *still* does not believe that he understands Chinese, no matter what the native Chinese speaker says. What I have in mind here is the following sort of situation: as it happens, I understand French to a certain extent; let us say that I believe that I understand 80% of what I hear or read, and that I can express myself with, say, 75% expressiveness: I can carry on a conversation on any topic (even give directions to

* Given that they were also taken in (perhaps) by the Great Oz himself, arguably they are not experts, but one can easily imagine a slightly different situation in which they would be. On the other hand, who's to say that, *from their point of view*, Oz was *not* a wizard?

Parisian taxi drivers), but I always feel that I am missing something or cannot quite generate the right idioms. Suppose, however, that a native French speaker tells me that I am fluent in French. "Ah, if only that were true," I reply. Who is right? Searle (in or out of the room) would say that *I* am – I do *not* (fully) understand French, no matter what the native French speaker tells me.

But Searle-in-the-room is not quite in my situation. He has the advantage of an instruction book (his Chinese natural-language-understanding and -generating program). And this suggests (as an anonymous reviewer pointed out) that our whole description of the Chinese Room is slightly misleading. Is it Searle-in-the-room with whom the native Chinese speaker is conversing? Or is it Searle-in-the-room together with his instruction book? Interestingly, it is quite clear that Searle himself, in his 1980 paper, assumes that it is Searle-in-the-room with whom the native Chinese speaker is conversing. There are, however, three candidates: The native Chinese speaker might be conversing with Searle-in-the-room, Searle-in-the-room + book, or the entire room (together with its contents). To see which it really should be (no matter whom Searle himself says it is), consider that the native Chinese speaker's interlocutor is supposed to be analogous to a computer running a natural-language-understanding and -generating program. We cannot align Searle-in-the-room (all by himself) with the computer, for the book (which must align with the program) is essential to the set-up. If we align the entire room with the computer, then Searle-in-the-room aligns with the central-processing unit, and the book aligns with the program.* If we align Searle-in-the-room + book to the computer, then the surrounding room is irrelevant (it plays the role of whatever in the Turing Test is used to hide the true nature of the interlocutors).

In all cases, it is not just Chineseless Searle-in-the-room who is conversing with the native Chinese speaker, but Searle-in-the-room + book. This is the "systems" reply to the Chinese-Room Argument (Searle, 1980: 419), and I am bringing it up for two reasons. First, it shows that, in the Chinese-Room situation, *unlike* my French situation, Searle-in-the-room by himself cannot insist that, because he (alone) knows no Chinese, his point of view takes precedence – because he is *not* alone: He has his instruction book, and, with its help, he does pass the Chinese-understanding test with flying colors, as judged by the only qualified judge there is. Were Searle-in-the-room, *with his book*, to be stranded on a desert island and forced to communicate with a Friday who only spoke Chinese, he – with the help of his book – would be able to do it. The native Chinese speaker is the only person qualified to say, truthfully, "I am conversing with someone who (or something that)

* Harnad (2000: §17) suggests that it is "spooky" to think that Searle-in-the-room does not understand Chinese but that the room including him does. But imagine a native Chinese speaker's brain (which aligns with Searle-in-the-room or with the CPU of a Chinese natural-language processor) saying to us, "Sorry; I don't know what you're talking about when you ask whether I 'understand Chinese.' I just fire neurons; some have pretty patterns (like non-programmatic music), but what does that have to do with understanding Chinese?" Searle-in-the-room can protest similarly. But clearly what the native Chinese speaker's brain is doing (and what Searle-in-the-room is doing) is essential to understanding Chinese.

understands Chinese." That someone (or something) has *no* right to assert that he (or she, or it) either does *or* does not speak Chinese.*

The second point to notice about the systems reply (although it is secondary to my present purpose) is that it is reminiscent of Hutchins's theory of "cognition in the wild" (Hutchins, 1995ab). The extended cognitive system that navigates a ship, consisting of the crew *plus* various instruments, is a real-life counterpart of Searle-in-the-room + book. Hutchins argues that it is not any individual crew member who navigates the ship, but the crew + instruments-that-are-external-to-the-crew's-minds: "systems that are larger than an individual may have cognitive properties in their own right that cannot be reduced to the cognitive properties of individual persons" (Hutchins, 1995b: 266). Similarly, I argue with the systems reply that Searle-in-the-room + the-instruction-book-that-is-external-to-his-mind has the cognitive property of understanding Chinese and that this is not (therefore) a cognitive property of Searle-in-the-room by himself (which – interestingly – is consistent with Searle-in-the-room's protestations that he (alone) still does not understand Chinese). To repeat, Searle-in-the-room's point of view is not the one that counts.

6.2. No Direct Access

To return to an earlier point, external links of the sort that Searle believes necessary are not needed, because the cognitive agent has no direct access to external entities. Those are fighting words, so what do I mean by them? I mean, simply, that if I want to say that 'tree' refers to that tree over there, I can only do so by associating my internal word 'tree' with my internal representative of that tree over there. Let me spell this out in more detail: I see a tree over there, and – while pointing to it – I say, "That's what 'tree' refers to" (or, more simply, "That's a tree"; but cf. Percy (1975: 258–264) on the dangers of this formulation). But what do I see? I am directly aware of the following visual image: my hand pointing to a tree. The visual image of the pointing hand and the visual image of the pointed-to tree are all internal. I go up and touch the tree (how much closer to the external world could I get?). But now all I have is an internal tactile image of the tree. It is all internal. I only indirectly access the external tree. ("[I]t is not really the world which is known but the idea or symbol ..., while that which it symbolizes, the great wide world, gradually vanishes into Kant's unknowable noumenon" (Percy, 1975: 33).)

Why do I believe that visual (and other sensory) images are internal, that I have no direct access to the external world, or, better, that my access to the external world – for I do believe that we have such access! – is always mediated by internal representatives of it? I am convinced by the following simple experiments (versions of the argument from illusion): look at some distant object, such as a small light source about 10 feet away. Close your left eye; you still see the light. Now open

* Cf. my Korean-Room Argument and my example of a student who does not understand what greatest common divisors are but who can compute them, in Rapaport (1988b: §§4–5).

your left eye and close your right; you still see it. But are you seeing the same thing you were in the two previous cases? In one sense, presumably, the answer is "Yes:" You are seeing the same distal object – but only indirectly and as mediated by an intensional representative. In another sense – the one I am interested in – the answer is "No:" The two (intentional) objects directly seen by your two eyes are slightly different (different locations relative to other entities in your visual field; different shapes; in my own case, at times, slightly different colors). And how do I know that there are two objects? Because, by crossing my eyes, I can see both at once (and, in so doing, I can compare their different colors)! Since, by hypothesis, there are not *two* of them in the external world, the internal images and the external object are even numerically distinct. (There is even a third object: the apparently 3-dimensional one constructed by stereoscopic vision (cf. Julesz, 1971), which differs in shape and location from the other two. All are internal visual images – representations of the external object. And the stereoscopically constructed image is not identical with the external object, precisely because it is constructed by the "mind's eye.")

I am not a pure solipsist, merely a representative realist. There *is* an external world, and my internal images are directly caused by external objects. But *I* have (perhaps better: my mind has) no (direct) access to the external objects. Does anyone? Surely, you say, *you* could have access to both worlds. From this third-person point of view, you could have access to my brain and to the external world, and – in the golden age of neuroscience – will be able to associate certain nerve firings with specific external objects. Similarly, I – as Cassie's programmer – can associate nodes of her semantic-network "mind" with things in the external world. Or consider again the situation in which I point to a tree and say 'tree.' From your point of view, you see both the tree and me pointing to it – both of which are, apparently, in the external world. Aren't we both looking at the same tree?

Not really. For suppose I associate Cassie's node `B1` (which, let us say, she lexicalizes as 'tree') with that tree over there. What am I really doing? I'm associating *my internal representation of Cassie's node* with *my internal representation of the tree*. And this is all internal to *me*. In the case of my pointing to the tree, all *you* are seeing is the following internal image: my hand pointing to a tree. We can only *assume* that there is an external tree causally responsible for our two internal-to-ourselves tree-images. This is what the third-person point of view really amounts to. ("Kant was rightly impressed by the thought that if we ask whether we have a correct conception of the world, we cannot step entirely outside our actual conceptions and theories to as to compare them with a world that is not conceptualized at all, a bare 'whatever there is' " (Williams, 1998: 40).)

So, by merging internalized semantic markers with (internal) syntactic markers, the semantic enterprise of mapping meanings to symbols can be handled by syntactic symbol (or marker) manipulation, and, thus, syntax *can* suffice for the (first-person) semantic enterprise.

7. Syntactic Semantics: II – A Recursive Theory of Semantic Understanding

There is a second way to approach syntactic semantics. Semantics is concerned with two domains and one binary relation: (1) the domain of the syntactic markers, characterized by (syntactic) formation or inference rules – call this the syntactic domain; (2) the domain of the semantic interpretation, the domain of the entities that are the meanings (or semantic interpretations) of the syntactic entities – call this the semantic domain; and (3) a mapping between the syntactic and semantic domains – the semantic interpretation.

What is the purpose of a semantic interpretation of a syntactic domain? Typically, we use the semantic domain to understand the syntactic domain. If we understand one thing in terms of another, ideally that other must already be understood. The semantic domain, therefore, must ideally be antecedently understood. How? There are two ways to understand the semantic domain: we could turn around and treat *it* as a syntactic domain – as a domain of (uninterpreted) markers characterized syntactically – and then find some *third* domain to play the role of semantic interpretation for *it*. And so on, in what Smith (1987) has called a "correspondence continuum." At some point, this process must stop. Our understanding of the last domain in the sequence must be in terms of the domain itself.*

And the only way to understand a domain in terms of itself is syntactically; i.e., we understand it by being conversant with manipulating its markers: that is what syntactic understanding amounts to (cf. Rapaport, 1986b). To give the most obvious example, we understand a deductive system syntactically when we understand it proof-theoretically. On this recursive picture of understanding, semantic understanding is, in the final analysis – the base case of the recursion – syntactic understanding. (It is also possible that the correspondence continuum ends in a circle of domains, each of which is understood in terms of the next one in the cycle. In this case, our understanding of any domain in the circle must always be relative to our understanding of the other domains. In fact, we would be better off considering the cycle of domains as a single, large domain, understood syntactically. For details and further discussion, see Rapaport (1995).)

I understand the internal symbols of my own Mentalese language of thought syntactically. One could say that "mental terms" do not *mean*; they just *are* (shades of Gertrude Stein?). More precisely, they *interact*: I manipulate them according to certain (no doubt unconscious) rules. Cassie does the same with her nodes. The meaning of any node in her semantic network consists, essentially, of its relations to all the other nodes in the entire network, or, as it is often put, its meaning is its

* For the sake of clarity, let me provide an example. Jurafsky and Martin (2000: Ch. 14) offer the first-order predicate calculus (FOPC) as a meaning-representation language (i.e., semantic domain) for providing the semantics of natural language (*ipso facto* considered as a syntactic domain). They then treat FOPC as a syntactic domain, and offer a "semantics of FOPC" (pp. 516ff) in terms of a "database semantics," which, they point out, is, in turn, to be understood as representing the real world. They appear to assume that we understand the real world directly. (For a fuller discussion of the issues involved in this "model muddle," see Wartofsky (1966); Rapaport (1995, 1996: Ch. 2).)

location in the network (cf. Carnap, 1928; Quine, 1951; Quillian, 1967; Rapaport, 1988b). For some purposes, this may be too much and would need to be constrained to some suitable subnetwork (cf. Hill, 1994, 1995; in this way, we can come to learn dictionary-like meanings of new words from context, without any recourse to external sources – cf. Ehrlich and Rapaport, 1997; Rapaport and Ehrlich, 2000).

How does Searle-in-the-room + book understand the native Chinese speaker? In the same way that I understand you: By mapping internal representations of your utterances, considered as syntactic entities, to my internal symbols (which, as we have seen, will include internal representations of external objects), and then doing symbol manipulation – syntax – on them. This is what Searle-in-the-room does: He maps internal representations of the native Chinese speaker's utterances (i.e., he maps the squiggle-input) to his internal symbols (as specified in the instruction book, which must – although Searle did not specify it – contain a knowledge-representation and reasoning system; cf. §8, below), and then he manipulates the symbols (see Rapaport, 1988b: §3.5.)

Here is where the two approaches to syntactic semantics merge. On the first view of syntactic semantics, the domain of interpretation of a syntactic system is "internalized" – converted into (more) syntactic markers – so that the semantic relations *between* the syntactic system and the semantic domain become syntactic relations *among* the markers of a (larger) syntactic system. On the second view of syntactic semantics, semantic interpretation is seen to be a recursive phenomenon whose base case is a (syntactic) system that can only be understood in terms of itself, i.e., in terms of the relations among its markers. Where the syntactic system is Mentalese, we find that there are two subsystems: There is a system of mental terms (the "old" markers) whose meanings are just the mental terms in the other subsystem (namely, the internalized representatives of the external semantic domain). And the system of those internalized representatives is understood syntactically. But, of course, the whole system consisting of both sorts of markers is just understood syntactically.*

As one anonymous reviewer noted, Searle could object that

> the reason why a word gets its meaning by being associated with a representation, is that ... it is associated with ... a *representation*, i.e. something which is somehow related to something external. Thus the link to the outer world is crucial after all, although it is now present in a disguise.

However, as I intimated before (Section 3.2.2; cf. Section 8, below, and Rapaport, 1988b), the links are merely causal links providing the internal markers that happen to be representatives of their causal origins. But, consistently with methodological solipsism (Fodor, 1980), we need not consider these causal histories when trying

* Where do the internal representatives – the initial concepts – come from? Each heard word is accompanied by a "bare-particular" concept (see Shapiro and Rapaport, 1995), whose only "content" is that it is that which is expressed by that word (cf. the semantics of the SNePS "lex" arc, Shapiro and Rapaport 1987). Connections to other concepts give it more detail. Thus, all such information is "assertional," not "structural," to use Woods's (1975) distinction.

to explain the semantic role of these markers. (For details, see Rapaport (1996: Ch. 6).)

8. The Mind as Syntactic System: A Game with Rules?

What is required of a cognitive agent for it to be able to understand and generate language in this syntactic fashion? A lot. It is not enough (as one anonymous reviewer suggested) for a computational cognitive agent to be endowed with "a list of all meaningful conversations shorter than a length so huge that no human can keep up a conversation for such a long time," along with a table-lookup program for this list (cf. Section 2.1, above). Such a computational cognitive agent would not be able to pass a Turing test, much less think, for no such list could possibly be complete: there is no way to predict in advance what the interrogator might ask it or what neologisms the interrogator might use (cf. Rapaport and Ehrlich, 2000), nor could it learn.

As I have urged before (Rapaport, 1988b, 1995), a computational cognitive agent will need to be able to do many things: take discourse (not just individual sentences) as input; understand all input, grammatical or not; perform inference and revise beliefs; make plans (including planning speech acts for natural-language generation, planning for asking and answering questions, and planning to initiate conversations); understand plans (including the speech-act plans of interlocutors); construct a "user model" of its interlocutor; learn (about the world and about language); have lots of knowledge (background knowledge; world knowledge; commonsense knowledge; and practical, "how-to," knowledge – see Erion, 2000); and remember what it heard before, what it learns, what it infers, and what beliefs it revised ("Oh yes, I used to believe that, but I don't any more"). And it must have effector organs to be able to *generate* language. In short, it must have a mind.* But note that the necessary mind, thus characterized, will be a purely syntactic system: a system of markers (perhaps semantic-network nodes, perhaps a neural network) and algorithms for manipulating them.

Such algorithms and markers are sometimes called "rules and representations," but I dislike that phrase. First, "rules" suggests rigid, unbreakable, antecedently-set-up laws. But the algorithms** for manipulating the markers need not be lawlike (they would probably need to be non-monotonic "default" or "defeasible" "rules"), and they could be created on the fly (the system has to be able to learn). Second,

* As Shapiro has pointed out to me, without the effectors, it might have a mind, but not one that would be detectable via a (purely linguistic) Turing test. Cf. the comments in Shapiro (1995: 521–522) concerning the cognitive abilities of humans with physical disabilities (see also Maloney, 1987: 352–353).

** Shapiro (1997) prefers the term "procedure" to "algorithm" because, on the standard introduction-to-computer-science definition of "algorithm," algorithms halt and are correct, but many interactive computational procedures (e.g., those for natural-language understanding and generation, or even an airline reservation system) do neither. See Rapaport (1998) for further discussion of what an algorithm is.

as I urged in Sections 4 and 7, the markers should not be thought of as symbols *representing* something external to the system; although they *can* be related to other things by a third person, the only relations needed by the cognitive agent are all internal. Finally, "rules and representations" is usually taken as a euphemism for what Haugeland (1985) called "GOFAI:" good old-fashioned, classical, symbolic AI (and often for a particular subspecies of GOFAI: production systems). But "markers and algorithms" applies equally well to connectionist, artificial neural networks, which disdain rules and representations as being too inflexible or too high-level, and everything that I have said about syntactic semantics applies to connectionist, artificial neural networks, taking the nodes of an artificial neural network as the markers.

9. Who Can Pass a Turing Test?

I believe that a suitably programmed computer could pass a Turing test. I do not think that this has happened yet, examples such as Eliza, Parry, or the Loebner competitions notwithstanding.* Nor do I think that it is going to happen in the near future. As I write, 2001 is close upon us, but HAL is not (cf. Stork, 1997), and I will not venture to make any more precise predictions: both Turing (who, in 1950, predicted 2000) and Simon and Newell (who, in 1957, predicted 1967 for the chess version of a Turing test, missing by 30 years; see Simon and Newell, 1958) were way off, and I could not hope to compete with the likes of them.**

But I believe that a suitably programmed computer *will*, eventually, pass a Turing test. And, more importantly, I believe that such a Turing-test-passing computer *will* "really" think, for the reasons adumbrated above, namely, syntax suffices for semantic understanding. More cautiously, I believe that it is a worthy research program to try to build such a computer (i.e., to write such programs) and that such an attempt is the *only* way to find out whether such a computer can be built (cf. Rapaport, 1998).

But there is another reason that a Turing test will eventually be passed. It is less interesting from a computational point of view, more so from a sociological point of view. It is simply that – to return to the earlier discussion of the Internet dog – for whatever reasons (and what these are is worth exploring), humans tend to treat other entities with which they interact as if they were human:

> As [software] agents are better able to create the illusion of artificial life, the social bond formed between agents, and the humans interacting with them, will

* On Eliza, see, e.g., Weizenbaum (1966). On Parry, see, e.g., Colby et al. (1972). On the Loebner competitions, see Loebner (1994), Shieber (1994a, 1994b).

** Although Simon says that "it had nothing to do with the Turing Test" and that "(a) I regard the predictions as a highly successful exercise in futurology, and (b) placed in the equivalent position today, I would make them again, and for the same reasons. (Some people never seem to learn.)" (personal communication, 24 September 1998). At the end of the next millennium, no doubt, historians looking back will find the 40-year distance between the time of Newell and Simon's prediction and the time of Kasparov's defeat to have been insignificant.

> grow stronger. New ethical questions arise. Each time we inspire an agent with one or more lifelike qualities, we muddy the distinction between users being amused, or assisted, by an unusual piece of software and users creating an emotional attachment of some kind with the embodied image that the lifeless agent projects (Elliott and Brzezinski, 1998: 15).

Call this "anthropomorphism" if you wish. Call it "intentional stance," if you prefer (Dennett, 1971). We have already witnessed tendencies along these lines with Eliza, the winners of the Loebner competitions, and even Kasparov's attitude toward Deep Blue.*

What will happen when we accept a computer as having passed a Turing test? Surely, I predict, we will accept it as a thinking thing. If that means, to paraphrase Turing, that the use of the word 'think' will have *altered* (or been metaphorically extended) "so much that one will be able to speak of machines thinking without expecting to be contradicted," so be it. But my main point in this paper has been to show that no such change is needed. "General educated opinion" will come to see that syntax suffices for *real* thinking.

Acknowledgements

I am grateful to Stuart C. Shapiro and the other members of the SNePS Research Group and to John T. Kearns, Justin Leiber, K. Nicholas Leibovic, Alan H. Lockwood, Ausonio Marras, James McGilvray, Paul E. Ritty, Elaine T. Stathopoulos, Susan B. Udin, and two anonymous reviewers for comments on earlier versions of this essay. An earlier version was presented as a master class in the Program in Philosophy and Computers and Cognitive Science (PACCS), SUNY Binghamton.

References

Abelson, R.P., 1968, "Simulation of social behavior," pp. 274–356 in *The Handbook of Social Psychology*, Vol. 2, 2nd edn., G. Lindzey and E. Aronson, eds., Reading, MA: Addison-Wesley.

Ackerman, D., 1989, "Penguins," *The New Yorker*, 10 July, 38–67.

Associated Press, 1997, "Opponent bytes at offer for draw with Kasparov," *Buffalo News*, 7 May, A7.

Baum, L.F., 1900, *The Wonderful Wizard of Oz*, New York: Dover (1966 reprint).

Berman, J. and Bruckman, A., 1999, "The Turing game," http://www.cc.gatech.edu/elc/turing/

Bunn, J.H., forthcoming, "Universal grammar or common syntax? A critical study of Jackendoff's *Patterns in the Mind*," *Minds and Machines*.

Carnap, R., 1928, *The Logical Structure of the World*, R.A. George (trans.), Berkeley, CA: University of California Press (1967).

Colby, K.M., Hilf, F.D., Weber, S., and Kraemer, H.C., 1972, "Turing-like indistinguishability tests for the validation of a computer simulation of paranoid processes," *Artificial Intelligence* **3**, 199–221.

Damasio, A.R., 1989, "Time-locked multiregional retroactivation," *Cognition* **33**, 25–62.

* Kasparov has spoken of Deep Blue using "intentional-stance" terminology: "Today I didn't play well, but the computer was a computer and eventually it *knew* enough not to lose the game" (Associated Press, 1997; my italics). Cf. Schonberg (1989: B2), Johnson (1997) and Levy (1997: 54).

Dennett, D.C., 1971, "Intentional systems," *Journal of Philosophy* **68**, 87–106.

Ehrlich, K. and Rapaport, W.J., 1997, "A computational theory of vocabulary expansion," pp. 205–210 in *Proceedings of the 19th Annual Conference of the Cognitive Science Society*, Mahwah, NJ: Erlbaum.

Elliott, C. and Brzezinski, J., 1998, "Autonomous agents as synthetic characters," *AI Magazine* **19**, 13–30.

Erion, G.J., 2000, "Common sense: An investigation in ontology, epistemology, and moral philosophy," Ph.D. Diss., Philosophy Department, SUNY Buffalo.

Fetzer, J.H., 1994, "Mental algorithms: Are minds computational systems?," *Pragmatics and Cognition* **2**, 1–29.

Fodor, J.A., 1980, "Methodological solipsism considered as a research strategy in cognitive psychology," *Behavioral and Brain Science* **3**, 63–109.

Hafner, K., 1999, "Guessing who is online," *The New York Times*, July 22.

Harnad, S., 2000, "Minds, machines and Turing: The indistinguishability of indistinguishables," *Journal of Logic, Language, and Information* **9**, this issue.

Haugeland, J., 1985, *Artificial Intelligence: The Very Idea*, Cambridge, MA: MIT.

Hill, R.K., 1994, "Issues of semantics in a semantic-network representation of belief," Tech. Rep. 94-11, Buffalo: SUNY Buffalo Computer Science Department.

Hill, R.K., 1995, "Non-well-founded set theory and the circular semantics of semantic networks," pp. 375–386 in *Intelligent Systems: 3rd Golden West International Conference*, E.A. Yfantis, ed., Dordrecht: Kluwer Academic Publishers.

Hutchins, E., 1995a, *Cognition in the Wild*, Cambridge, MA: MIT.

Hutchins, E., 1995b, "How a cockpit remembers its speeds," *Cognitive Science* **19**, 265–288.

Jackendoff, R., forthcoming, "Bringing *Patterns* into focus: A response to Bunn," *Minds and Machines*.

Johnson, G., 1997, "Ghost in the chess machine: Brain or box? Think about it," *The New York Times*, 9 May, A1, B4.

Julesz, B., 1971, *Foundations of Cyclopean Perception*, Chicago, IL: University of Chicago Press.

Jurafsky, D. and Martin, J.H., 2000, *Speech and Language Processing*, Englewood Cliffs, NJ: Prentice-Hall.

Kearns, J., 1997, "Propositional logic of supposition and assertion," *Notre Dame Journal of Formal Logic* **38**, 325–349.

Lassègue, J., 1996, 'What kind of Turing test did Turing have in mind?," *Tekhnema: Journal of Philosophy and Technology* **3**, http://www.gold.ac.uk/tekhnema/3/lassegue/read01.html

Levy, S., 1997, "Man v. machine," *Newsweek*, 5 May, 51–56.

Loebner, H.G., 1994, "In response [to Shieber 1994a]," *CACM* **37**(6), 79–82.

Maida, A.S. and Shapiro, S.C., 1982, "Intensional concepts in propositional semantic networks," *Cognitive Science* **6**, 291–330.

Maloney, J.C., 1987, "The right stuff," *Synthese* **70**, 349–372.

Manin, Yu.I., 1977, *A Course in Mathematical Logic*, New York: Springer-Verlag.

McGilvray, J., 1998, "Meanings are syntactically individuated and found in the head," *Mind and Language* **13**, 225–280.

Morris, C., 1938, *Foundations of the Theory of Signs*, Chicago, IL: Unversity of Chicago Press.

Percy, W., 1975, *The Message in the Bottle*, New York: Farrar, Straus and Giroux.

Quillian, M.R., 1967, "Word concepts: A theory and simulation of some basic semantic capabilities," *Behavioral Science* **12**, 410–430.

Quine, W.V.O., 1951, "Two dogmas of empiricism," reprinted in *From a Logical Point of View*, 2nd edn., Rev., Cambridge, MA: Harvard University Press, 1980, pp. 20–46.

Rapaport, W.J., 1978, "Meinongian theories and a Russellian paradox," *Noûs* **12**, 153–180; errata, 1979, *Noûs* **13**, 125.

Rapaport, W.J., 1981, "How to make the world fit our language: An essay in Meinongian semantics," *Grazer Philosophische Studien* **14**, 1–21.

Rapaport, W.J., 1985, "Machine understanding and data abstraction in Searle's Chinese Room," pp. 341–345 in *Proceedings of the 7th Annual Conference of the Cognitive Science Society*, Hillsdale, NJ: Erlbaum.

Rapaport, W.J., 1986a, "Philosophy, artificial intelligence, and the Chinese-Room Argument," *Abacus* **3**, Summer, 6–17; correspondence, 1987, *Abacus* **4**, Winter, 6–7, *Abacus* **4**, Spring, 5–7.

Rapaport, W.J., 1986b, "Searle's experiments with thought," *Philosophy of Science* **53**, 271–279.

Rapaport, W.J., 1988a, "To think or not to think," *Noûs* **22**, 585–609.

Rapaport, W.J., 1988b, "Syntactic semantics: Foundations of computational natural-language understanding," pp. 81–131 in *Aspects of Artificial Intelligence*, J.H. Fetzer, ed., Dordrecht: Kluwer Academic Publishers.

Rapaport, W.J., 1995, "Understanding understanding: Syntactic semantics and computational cognition," pp. 49–88 in *AI, Connectionism, and Philosophical Psychology*, J.E. Tomberlin, ed., Phil. Perspectives, Vol. 9, Atascadero, CA: Ridgeview.

Rapaport, W.J., 1996, *Understanding Understanding: Semantics, Computation, and Cognition*, Tech. Rep. 96-26, Buffalo: SUNY Buffalo Computer Science Department; http://www.cse.buffalo.edu/tech-reports/96-26.ps.Z

Rapaport, W.J., 1998, "How minds can be computational systems," *Journal of Experimental, Theoretical and Artificial Intelligence* **10**, 403–419.

Rapaport, W.J., 1999, "Implementation is semantic interpretation," *The Monist* **82**, 109–130.

Rapaport, W.J. and Ehrlich, K., 2000, "A computational theory of vocabulary acquisition," in *Natural Language Processing and Knowledge Representation*, Ł. Iwańska and S.C. Shapiro, eds., Menlo Park, CA/Cambridge, MA: AAAI/MIT (in press).

Schonberg, H.C., 1989, "Kasparov beats chess computer (for now)," *New York Times*, 23 October, A1, B2.

Searle, J.R., 1980, "Minds, brains, and programs," *Behavioral and Brain Science* **3**, 417–457.

Searle, J.R., 1982, "The myth of the computer," *New York Review of Books*, 29 April, 3–6; cf. correspondence, same journal, 24 June 1982, 56–57.

Shapiro, S.C., 1979, "The SNePS semantic network processing system," pp. 179–203 in *Associative Networks*, N. Findler, ed., New York: Academic Press.

Shapiro, S.C., 1995, "Computationalism," *Minds and Machines* **5**, 517–524.

Shapiro, S.C., 1997, "What is computer science?," http://www.cse.buffalo.edu/~shapiro/Papers/whatiscs.pdf

Shapiro, S.C., 1998, "Embodied Cassie," pp. 136–143 in *Cognitive Robotics: Papers from the 1998 AAAI Fall Symposium*, Tech. Rep. FS-98-02, Menlo Park, CA: AAAI.

Shapiro, S.C. and Rapaport, W.J., 1987, "SNePS considered as a fully intensional propositional semantic network," pp. 262–315 in *The Knowledge Frontier*, N. Cercone and G. McCalla, eds., New York: Springer-Verlag.

Shapiro, S.C. and Rapaport, W.J., 1991, "Models and minds: Knowledge representation for natural-language competence," pp. 215–259 in *Philosophy and AI*, R. Cummins and J. Pollock, eds., Cambridge, MA: MIT.

Shapiro, S.C. and Rapaport, W.J., 1992, "The SNePS family," *Computers and Mathematics with Applications* **23**, 243–275.

Shapiro, S.C. and Rapaport, W.J., 1995, "An introduction to a computational reader of narrative," pp. 79–105 in *Deixis in Narrative*, J.F. Duchan, G.A. Bruder, and L.E. Hewitt, eds., Hillsdale, NJ: Erlbaum.

Shieber, S.M., 1994a, "Lessons from a restricted Turing test," *CACM* **37**(6), 70–78.

Shieber, S.M., 1994b, "On Loebner's lessons," *CACM* **37**, 83–84.

Simon, H.A. and Newell, A., 1958, "Heuristic problem solving: The next advance in operations research," *Operations Research* **6**(6), 1–10.

Simpson, J.A. and Weiner, E.S.C. (preparers), 1989, *The Oxford English Dictionary*, 2nd edn., Oxford: Clarendon.

Smith, B.C., 1987, "The correspondence continuum," Report CSLI-87-71, Stanford, CA: CSLI.

Srihari, R.K., 1991, "PICTION: A system that uses captions to label human faces in newspaper photographs," pp. 80–85 in *Proceedings of the 9th National Conference on Artificial Intelligence (AAAI-91)*, Menlo Park, CA: AAAI/MIT.

Srihari, R.K. and Rapaport, W.J., 1989, "Extracting visual information from text: Using captions to label human faces in newspaper photographs," pp. 364–371 in *Proceedings of the 11th Annual Conference of the Cognitive Science Society*, Hillsdale, NJ: Erlbaum.

Stork, D.G., 1997, *HAL's Legacy*, Cambridge, MA: MIT.

Thagard, P., 1986, "Parallel computation and the mind-body problem," *Cognitive Science* **10**, 301–318.

Turing, A.M., 1936, "On computable numbers, with an application to the *Entscheidungsproblem*"; reprinted, with corrections, 1965 in *The Undecidable*, M. Davis, ed., New York: Raven, pp. 116–154.

Turing, A.M., 1950, "Computing machinery and intelligence," *Mind* **59**, 433–460.

Wartofsky, M.W., 1966, "The model muddle," pp. 1–11 in *Models: Representation and the Scientific Understanding*, Dordrecht: Reidel (1979).

Weizenbaum, J., 1966, "ELIZA – A computer program for the study of natural language communication between man and machine," *CACM* **9**, 36–45.

Williams, B., 1998, "The end of explanation?," *The New York Review of Books* **45**, 40–44 (19 November).

Woods, W.A., 1975, "What's in a link," pp. 35–82 in *Representation and Understanding*, D.G. Bobrow and A.M. Collins, eds., New York: Academic Press.

View publication stats